LIVING BEYOND OUR MEANS

ROGER KOENKER

1. Introduction

As someone who has been advocating "living beyond one's means" since before the Reagan administration, I welcome Thomas's endorsement of the idea. If we embrace the maxim that statistics is about *variation* then we should stand ready to explore how that variation varies with covariates. Assuming, as theory may tempt us to do, that covariates shift only the central tendency of the response, while variation around the central tendency remains unperturbed is rarely plausible. Signal plus iid noise is a dangerous fiction.

Nonparametric regression methods have encouraged us to think much more flexibly about how covariate effects vary over x-space, slicing design space into segments that each deserve their own local estimate. Quantile regression seeks to do something similar, slicing the range of the response variable into local conditional quantile functions. Combining the two approaches by estimating nonparametric conditional quantile functions offers a flexible way of characterizing the entire conditional distribution of $Y|X$.

While I find myself completely in accord with the *ends* set forth in "Beyond," I cannot condone many of the *means* employed to achieve them. While not descending to the depths of the Cheneyesque – rendition, black ops, waterboarding – there is an element of (data) torture that risks undermining the whole enterprise. I'll begin by explaining some of my reservations about parametric alternatives and their Bayesian elaborations for nonparametric quantile regression. To conclude there will be some malicious comments about expectiles.

2. The Paranormal and the Parametric

There is a strong tradition in statistics, going back at least to Karl Pearson, of what might be called distributional hybridization, grafting new parametric models onto old root stock. This can be very successful, we need look no further than the Box-Cox transformation, but it rarely proves to be the panacea enthusiasts claim. The attraction of global parametric models and their associated likelihood based estimation and inference methods is undeniable, but these siren songs have their

Version: February 18, 2014. This research was partially supported by NSF grant SES-11-53548. These comments were prepared as part of the discussion of the paper Beyond Mean Regression by Thomas Kneib, and appears in the journal *Statistical Modelling* in 2013.

dangers and it may be safer to follow the advice of Circe to plug our ears with beeswax, tie ourselves to the mast and sail nonparametrically on by.

Quantile regression was intended to be an exploratory data analysis tool. Often it enables us to see things that conventional parametric models render obscure, or even invisible. D. R. Cox concluding his discussion of a paper on Bayesian inference read to the Royal Statistical Society by Jimmy Savage wrote:

> A final general comment is that the discussion above is of the question of how to reach conclusions about parameters in a model on which we are agreed. It seems to me, however, that a more important matter is how to formulate more realistic models that will enable scientifically more searching questions to be asked of data. (Cox (1962))

Once we have taken a broader view we can always try to construct a narrower parametric view of the same material, hopefully one that doesn't do too much violence to what we first saw. But without the initial, exploratory look it is hard to be confident of the second stage.

Thus, I have nothing against **gamlss** as long as one is aware of its limitations: the inherent non-robustness due to sensitivity of moment-based estimators to tail behavior, the difficulty of interpreting higher order moment parameters and the difficulties of effectively linking the dependence of these parameters to observable covariates. These difficulties were all brought home to me while working on Wei, Pere, Koenker, and He (2005) in which we compared the elegant Cole and Green (1992) parametric Box-Cox procedure for estimating reference growth charts with a nonparametric quantile regression procedure. The Cole-Green method employed conventional $\mathcal{L}_2$ smoothing penalties on the second derivatives of the parameters $\{\mu(t), \sigma(t), \lambda(t)\}$ that were assumed to characterize the age specific distribution of childrens' heights. With considerable fine tuning we managed to adjust the smoothing parameters of this model to achieve growth charts that resembled the nonparametric charts we had estimated with conventional quantile regression on a few B-spline basis functions, but only at the cost of highly variable estimates of the Box-Cox $\lambda(t)$ trajectory that seemed highly implausible. In more complicated settings with more covariates these difficulties would be compounded.

3. Bayes in the Baño

Recent interest in Bayesian nonparametric methods has spawned considerable work on various forms of Bayesian quantile regression. The usual knock about Bayesian methods focuses on the difficulty of coming up with sensible priors. I've never quite understood this complaint; of course it isn't easy especially in high dimensional problems, anyone who thinks it is should consult the recent exchange between Larry Wasserman and Chris Sims. But everyone is entitled to the courage of their own convictions provided that they are not too dogmatic. It is just this last proviso that really worries me about the other crucial ingredient of the Bayesian paradigm: how is it that one can be so ignorant about model parameters but so confident about the specification of the likelihood? Likelihoods are especially problematic for quantile regression: if you are brave, or foolhardy, enough to write

$$Q_Y(\tau|x) = x^\top \beta(\tau) \qquad \tau \in (0,1),$$

then implicitly you have written a global model with log likelihood,

$$\ell(\beta|y,x) = -\sum \log(x_i^\top \dot{\beta}(\tau_i))$$

where $\dot{\beta}(\tau) = (d\beta_j(\tau)/d\tau)$. But this is not especially tractible, one would need *inter alia* to replace the mysterious, τ_i's by something like $F_Y(y_i|x_i)$, the conditional distribution of Y evaluated at the observed y_i's, or less hypothetically, by an estimate thereof. This parametric dilemma has led to several more ad hoc proposals, many of which fall back upon the asymmetric Laplace distribution, or ALD, model. As fond as I might be of the logarithm of this density, it doesn't serve very well as a likelihood. Of course, it does when logged bring us back to the familiar quantile regression objective function, but no one ever intended this to be a plausible *global* description of how data was generated. The whole point is that it is only a local description. It was simply a convenient gimmick that delivered the sample quantiles in univariate samples, and miraculously gave estimates of conditional quantile functions in various regression settings.

It is sometimes claimed that Bayesian posteriors based on such likelihoods deliver, as foretold by the Bernstein-von-Mises theorem, consistent estimates. This is true insofar as we are willing to accept the multitude of consistency results for quantile regression in its original form, but as soon as one would like to do inference, construct confidence, or credibility regions, all bets are off. And as we know: a Bayesian without betting, is like a monk without praying. We might at least consider that such a likelihood should have a free scale parameter, but once we start down this road there is no stopping. The parametric quagmire deepens and we sink slowly into it, left wondering if we could pull ourselves out by our hair, or our bootstraps if they aren't already submerged.

An interesting alternative to this ALD morass is a recent proposal of Reich, Fuentes, and Dunson (2011), who suggest in effect to replace the data by the point estimates of the quantile regression process, $\{\beta(\tau) : \ \tau \in (0,1)\}$, and to use the (estimated! asymptotic!) distribution of $\hat{\beta}(\tau)$ as a likelihood. Remarkably, this proposal can be traced back to one of the sacred texts of Bayesianism, see Jeffreys (1939). Check the index for "median." It seems safe to bet that this will remain a Bayesian heresy well into the next millenium.

My reservations about Bayesian likelihoods for quantile regression notwithstanding, penalty methods for disciplining the at times unruly behavior of quantile regression estimates are quite indispensable. Call them priors if you wish, or just invoke Tukey's "borrowing strength" dictum, and let 1000 of these flowers bloom. My personal preference leans toward the total variation penalties of Koenker, Ng, and Portnoy (1994) and Koenker and Mizera (2004), which are simply $\mathcal{L}_1$ variants of the more familiar Wabha-esque, $\mathcal{L}_2$ penalties appearing in "Beyond." The TV penalties yield piecewise linear fitted functions and have the advantage that it is relatively easy to incorporate shape constraints like monotonicity and/or convexity/concavity. Additive models of the type advocated in "Beyond" can be estimated using quantile fidelity and total variation penalties with the `rqss( )` in my **quantreg** R package. This framework keeps us well within the briar patch of linear programming, which when one is careful to exploit the inherent sparsity of the linear algebra is computationally very efficient.

There is always a lingering concern that the local nature of quantile regression estimation may produce non-monotone estimates of the conditional quantile function at some design points. This is the price we pay for the flexibility of local fitting; there is no free lunch. Global models, in contrast, deliver happy families of non-crossing quantile functions at the cost of imposing a potentially overly restrictive structure. Fortunately, there are several excellent remedies for non-monotonicity when it arises. Dette and Volgushev (2008) and Chernozhukov, Fernàndez-Val, and Galichon (2010) have shown that rearrangement can be employed without altering the asymptotic behavior of the quantile regression process, while improving its higher order bias properties. More traditional monotonization methods such as pool adjacent violators can be used similarly, and joint estimation of quantile regression parameters over several τ's has also be explored in several recent papers. Of course, in extreme circumstances where crossing of estimated conditional quantile functions is severe, one should seriously reconsider the specification of the model, often the linearity assumption can be highly suspect. In such cases non-monotonic estimates serve a valuable diagnostic purpose.

For model selection the analogue of the TV penalty is the ubiquitous "lasso" penalty which has proven to be an inspiration for countless theoretical papers, and – perhaps to a lesser degree – a practical tool for applied data analysis. "Beyond" favors boosting over the lasso and I would be very interested to learn more about why? A principle that I try to drum into students is: Any worthwhile estimate deserves an estimated measure of precision. This is difficult advice to adhere to, especially in high dimensional nonparametric problems. It was only after I had the opportunity to read the marvelous paper by Krivobokova, Kneib, and Claeskens (2010) that I felt confident enough to listen to my own advice, and – following their remarkable lead for additive models for the conditional mean – try to provide similar inference methods for additive modeling of quantile regression in `rqss( )`. The results of this adventure are reported in Koenker (2011), and serve to reenforce the encouraging message of earlier work about the effectiveness of Hotelling tube methods for constructing uniform nonparametric confidence bands. To my knowledge there is no such commensurate approach for boosting and I would be curious to know how one might try to rationalize this sin of omission.

4. Expectiles and Expectoration

There is one final question: Do expectiles clear the bar defining obscenity set by the U.S. Supreme Court, and immortalized by Lenny Bruce: are they – or are they not – *utterly without any redeeming social value*? Its a tough call, but just for the sake of argument I'll take the view that expectiles are, indeed, UWARSV. Many aspects of the case against expectiles are familiar: they are slippery, although they seek to describe a local property of a distribution, they depend on global properties of that distribution; they are inherently nonrobust, by manipulating the tails of the distribution one can make the expectiles dance at your will; and they are not equivariant to monotone transformations as are the quantiles. I could rest my case here, but why? – when we are having fun.

In this spirit I suppose one could ask: Why would we ever want to square our errors? Doesn't this simply exaggerate the importance of our biggest modeling blunders? I can hear you thinking: "How could we do without the mean, and didn't we come to our beloved mean by squaring errors?" To this one can hardly object, but the cautionary results of Bahadur and Savage (1956) are always worth keeping in mind.

Over the years I've made several (wildly unsuccessful) attempts to dislodge the loogie of expectorate from the collective statistical throat. The first of these attempts appeared nearly invisibly as an ET problem, Koenker (1992), asking whether there was a distribution for which the quantiles and the expectiles coincided. Of course I knew, because ET required submitted answers to proposed problems, that such a distribution did exist and that its density took the form,

$$f(y) = 2|y|/[(4 + y^2)^2 \sqrt{1 - 4/(4 + y^2)}].$$

At the time I thought that the world would take one look at this, marvel at its absurdity and immediately forget about expectiles altogether. Ah, the naiveté of (relative) youth. Note that the density displayed above has algebraic tails with (Pareto) tail exponent 2, and consequently the distribution has no variance, so the standard asymptotic theory for expectiles fails totally for this special distribution, the only one for which interpretation of expectiles as quantiles has validity!

My second attempt at jousting at the expectile windmill appeared as Figure 2.10 of Koenker (2005) and is reproduced below. It illustrates in grey a family of conditional quantile functions for a model with linear cqfs for $\tau \leq 1/2$ and quadratic cqfs for $\tau > 1/2$. Superimposed on the plot in black are the corresponding family of expectiles for the same model, and it is apparent that they all exhibit a nonlinearity that they inherit from the upper tail of the conditional distribution. Can this be a Good Thing? Hardly. Did it slow the flow of expectorate? Hardly. It is sometimes claimed that the conditional expectiles can be relabeled to obtain the conditional quantile functions;

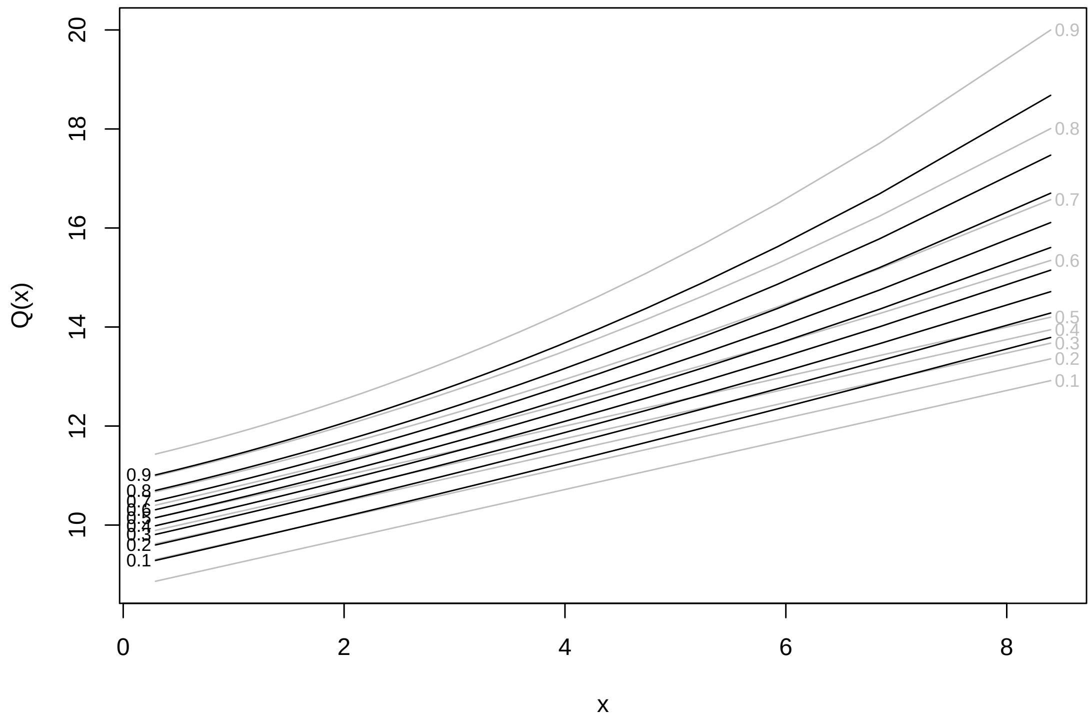

FIGURE 1. Conditional quantile functions and Conditional Expectile Functions: A family of conditional quantile functions (in grey) and the corresponding family of conditional expectile functions (in black). Note that the conditional quantile functions are linear in x, for $\tau \leq 1/2$, and quadratic for $\tau > 1/2$. However, due to the global nature of the expectiles the lower expectiles inherit the nonlinearity from the upper tail. The 0.10 expectile is roughly the same as the 0.20 quantile for x near 0, but it corresponds to the 0.35 quantile when $x = 10$.

this *cannot* be done in (generic) cases like the one we have illustrated. In the iid location-scale shift model, provided there is first moment, relabeling is possible, but I will argue that even in the simplest univariate cases this is a fool's errand.

A frequent justification of expectiles seems to be that they are "easy to compute," implying I suppose that they are easier to compute than quantiles. This is hard to dispute since it depends on so many unspecified factors. Easy for whom? For what problems? With what tools, what languages? In practice "easy" seems to translate as "programmable as iteratively reweighted least squares." As noted by

Schabel and Eilers (2009) $\mathcal{L}_2$ smoothing penalties can be easily incorporated via data augmentation, and this leads to a convenient smoothing approach for both univariate and bivariate additive components. While IWLS may be easy for the programmer, it is certainly not so easy for the poor machine. Modern quadratic programming methods like second order cone programming would be far more efficient. It would also enable investigators to consider much larger applications by exploiting sparse linear algebra. Of course quantile regression estimation with $\mathcal{L}_2$ smoothing penalties can also be formulated as SOCP just as well as expectiles, and the computational difficulty of the two problems are thus essentially identical.

"Beyond" offers one last claim for the superiority of expectiles over quantiles, a claim so preposterous that it probably needs no refutation, but just for the sake of completeness it seems imperative to offer some response. I quote,

> However, as Kauermann, Schulze-Waltrup, Sobotka, and Kneib (2012) have shown, expectiles can be easily transformed to calculate quantiles and may then also be more efficient.

Let's try to deconstruct this. First, we must be talking about univariate situations since as we have seen conditional expectile functions have different functional form than conditional quantiles so the transformation project must be conducted design point by design point. Second, even for univariate expectiles we will need to estimate many expectiles in order to infer something useful about any one quantile. Of course there are special cases: for symmetric distributions, F with density f the mean can be a more (asymptotically) efficient estimator of the median provided that the variance of the distribution is less than $1/(4f(F^{-1}(1/2)))$. This result was already established by Laplace, and was reprised by Edgeworth and Kolmogorov, but it is evident that the condition for an efficiency improvement is very special; at the normal model there is an asymptotic relative efficiency gain from using the mean of about one third, but if we encounter heavier tails the efficiency loss can be arbitrarily large. As soon as we move away from this symmetric setting there is no simple mapping from expectiles to quantiles, instead we have a linear operator that maps the whole expectile function back to the quantile function. In practice, what is proposed in Kauermann et. al. is to evaluate this operator on a grid. One can view this as a smoothing operation imposed on a modified version of the empirical distribution function.

The "easy" transformation from expectiles to quantiles developed in Kauermann, Schulze-Waltrup, Sobotka, and Kneib (2012) relies on a characterization of the τth expectile, say $\mu(\tau)$, of a random variable with distribution function, F as,

$$\mu(\tau) = \frac{(1-\tau)G(\mu(\tau)) + \tau(\mu(1/2) - G(\mu(\tau))}{(1-\tau)F(\mu(\tau)) + \tau(1 - F(\mu(\tau))} \tag{1}$$

where $G(\tau) = \int_0^\tau F^{-1}(u)du$. Given a vector of estimated expectiles, $\hat{\mu}(\tau_j): \ j = 1, \cdots, J$, for $0 < \tau_1 < \cdots < \tau_J < 1$, they write,

$$\hat{F}(\hat{\mu}(\tau_k)) = \sum_{j=1}^{k} \hat{\gamma}_j$$

and

$$\hat{G}(\hat{\mu}(\tau_k)) = \sum_{j=1}^{k} \hat{c}_j \hat{\gamma}_j$$

where $\hat{c}_j = (\hat{\mu}(\tau_j) + \hat{\mu}(\tau_{j-1}))/2$ for $j = 1, \cdots, J$. The grid of estimated expectiles is extended to define $\hat{\mu}(0) = X_{(1)}$ and $\hat{\mu}(1) = X_{(n)}$, the minimal and maximal order statistics of the sample, respectively. Substituting the foregoing expressions for $\hat{F}$ and $\hat{G}$ into (1), yields, with a little adjustment to delete an uninformative mean equation and add an equation in the upper tail, a *linear* system of J equations in the J unknown $\hat{\gamma}_j$'s. Given this surrogate estimate of F at the specified expectile grid points, linear interpolation is used to invert to obtain quantile estimates. In Figure 2 I've illustrated two estimates of the quantile function based on 199 standard Gaussian observations. The darker, more jagged curve is the expectile estimate using the Kauermann, Schulze-Waltrup, Sobotka, and Kneib (2012) procedure. As the figure illustrates these estimates of the quantile function are typically not monotone, but this can be rectified quite easily as we have discussed above. An even more disturbing aspect is the discrepancy one sees in the tails, which is caused by the fact that extreme Gaussian expectiles are actually much more central than the corresponding quantiles so there is typically a long interpolated segment in the tails. This effect is accentuated by using fewer expectiles, like the 25 suggested in Kauermann et al. The real question is: how is all this leading to an improvement over classical methods? Do we really expect an improvement over inverting the empirical distribution function? If so under what conditions?

To answer these questions let's briefly review some established theory. Pfanzagl (1976) has proven that:

> No translation equivariant and asymptotically uniformly median unbiased estimator is asymptotically more concentrated about the distribution quantile than the sample quantile.

Thus, we cannot expect improvements in the leading ($\mathcal{O}(n^{-1})$) term of the asymptotic expansion of the mean square error of estimators of the τth quantile over that achieved by the τth sample quantile. However, there is an extensive literature exploring the possibility of lower order improvements. Reiss (1980) considers quasi-quantiles of the form,

$$\tilde{Q}_m(\tau) = (X_{([n\tau]-m)} + X_{([n\tau]+m)})/2,$$

and shows that for appropriate choice of $m = m_n = \mathcal{O}(n^{2/3})$, one can ahieve MSE improvements of $\mathcal{O}(n^{-4/3})$, which are negligible relative to the $n^{-1}\tau(1-\tau)[Q'(\tau)]^2 =$

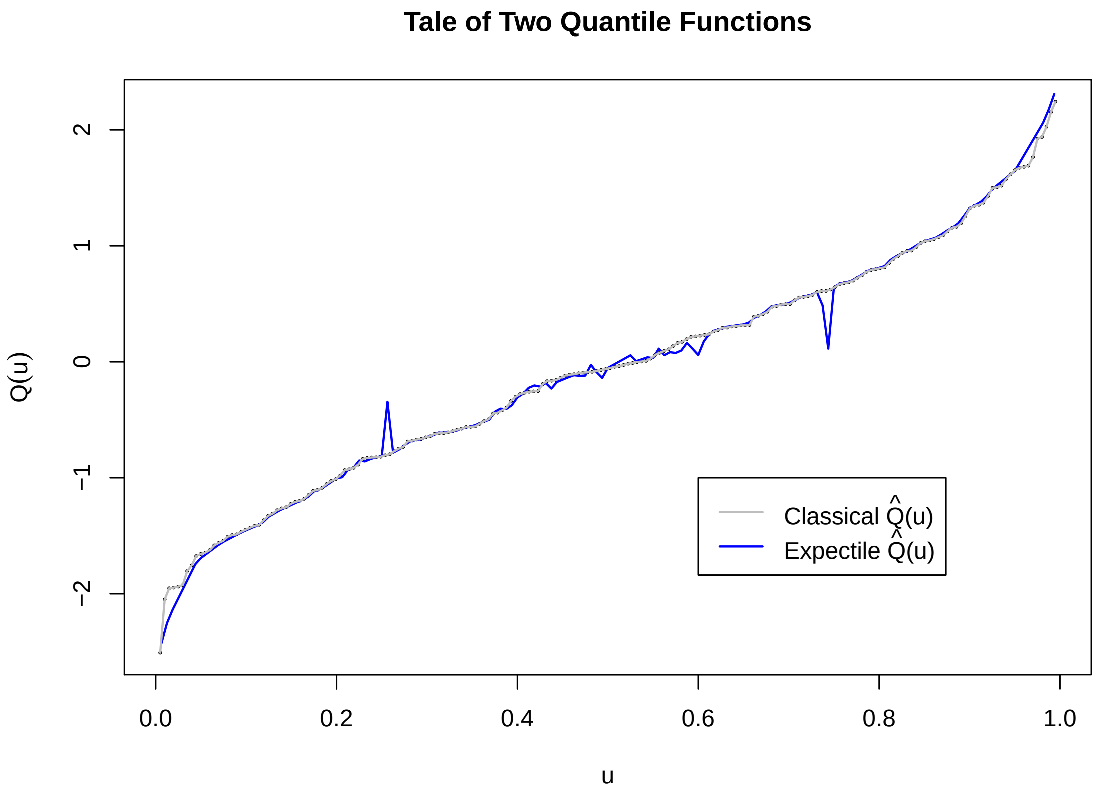

Figure 2. Two estimates of a standard Gaussian quantile function based on 199 observations: The grey curve depicts the classical quantile function based on inversion of the empirical distribution function. The darker, non-monotone curve is the estimate based on the Kauermann et al procedure of "inverting" the estimate of distribution function based on the expectiles. The latter curve is based on 160 equally spaced expectiles.

$\mathcal{O}(n^{-1})$ leading term, but nevertheless of interest. Parzen (1979), Falk (1984) and others have considered kernel smoothing of nearby order statistics as an alternative to these linear interpolents. Sheather and Marron (1990) provide an overview of this literature and show that for smooth quantile functions, with continuous $Q''(t)$ in a neighborhood of $t = \tau$, and kernel k, a compactly supported density symmetric about 0, the optimal bandwidth, h^* for such kernel estimates,

$$\tilde{Q}_h(\tau) = \sum_{i=1}^{n} h^{-1}\left(\int_{(i-1)/n}^{i/n} k((t-\tau)/h)dt\right)X_{(i)}$$

is $h^* = \alpha(k)\beta(Q)n^{-1/3}$, where

$$\alpha(k) = [2\int_{-\infty}^{\infty} uk(u)K(u)du/(\int_{-\infty}^{\infty} u^2 k(u)du)^2]^{1/3}$$

$K(u) = \int_{-\infty}^{u} k(u)du$ and $\beta(Q) = (Q'(\tau)/Q''(\tau))^{2/3}$. With this choice of bandwidth,

$$MSE(\tilde{Q}_h(\tau)) = n^{-1}\tau(1-\tau)[Q'(\tau)]^2 + \mathcal{O}(n^{-4/3}),$$

so again we have, at least potentially, some improvement over what is achievable with the ordinary sample quantile, but the improvement is asymptotically negligible. As is usual in such circumstances, it is handy to know about the local behavior of Q near τ, in particular the factor, $\beta(Q)$, in choosing the bandwidth, but based on their extensive simulations Sheather and Marron (1990) conclude that "even if one knew the best estimator [i.e. bandwidth] for each situation, one can expect an average improvement in efficiency of only 15%."

In Koenker (2005) such kernel smoothing methods are considered in the quantile regression context and a small simulation experiment is described that confirms the modest MSE gains seen in the univariate setting. Where does this leave the claim from "Beyond?" With careful attention to the monotonization of the Kauermann, Schulze-Waltrup, Sobotka, and Kneib (2012) procedure, it seems possible to reconstruct something close to the usual empirical distribution function, and thus with further careful choice of bandwidth one might be able to capture some of the benefits of these earlier smoothing proposals. But we cannot expect to improve upon the empirical distribution function, so if the modest improvement due to smoothing is deemed worthwhile, why not start there and dispense with the expectiles entirely?

I don't want to leave the impression that I am opposed to smoothing, or in the Tukey jargon "borrowing strength," in estimating either conditional or unconditional quantile models. Indeed, this can be a valuable way to increase the precision of estimates. In the center of the distribution these benefits are necessarily modest, but in the tails such smoothing is really essential. Recent work by Chernozhukov (2005) and Wang, Li, and He (2012) vividly illustrate this point. What I fail to see is any benefit derived from introducing the expectiles. Expectiles belong in the spittoon.

References

Bahadur, R. R., and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Ann. Math. Statist.*, 27, 1115–1122.

Chernozhukov, V. (2005): "Extremal Quantile Regression," *Annals of Statistics*, 33, 806–839.

Chernozhukov, V., I. Fernàndez-Val, and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Cole, T. J., and P. Green (1992): "Smoothing Reference Centile Curves: The LMS method and penalized likelihood," *Statistics in Medicine*, 11, 1305–1319.

Cox, D. R. (1962): "Comment on L.J. Savage's lecture "Subjective Probability and Statistical Practice"," in *The Foundations of Statistical Inference*, ed. by G. A. Barnard, and D. R. Cox. London: Methuen.

Dette, H., and S. Volgushev (2008): "Non-crossing non-parametric estimates of quantile curves," *J. Royal Stat. Soc.(B)*, 70, 609–627.

Falk, M. (1984): "Relative Deficiency of Kernel Type Estimators of Quantiles," *Ann. Statist.*, 12, 261–268.

Jeffreys, S. (1939): *Theory of probability*. Oxford University Press.

Kauermann, G., L. Schulze-Waltrup, F. Sobotka, and T. Kneib (2012): "Quantile or Expectile Regression: Which is Better?," Technical Report.

Koenker, R. (1992): "When are expectiles percentiles?," *Econometric Theory*, 8, 423.

——— (1993): "When are expectiles percentiles? Solution," *Econometric Theory*, 9, 526.

——— (2005): *Quantile Regression*. Cambridge U. Press.

——— (2011): "Additive models for quantile regression: Model selection and confidence bandaids," *Braz. J. Probab. Stat.*, 25, 239–262.

Koenker, R., and I. Mizera (2004): "Penalized triograms: total variation regularization for bivariate smoothing," *J. of Royal Stat. Soc (B)*, 66, 145–163.

Koenker, R., P. Ng, and S. Portnoy (1994): "Quantile smoothing splines," *Biometrika*, 81, 673–680.

Krivobokova, T., T. Kneib, and G. Claeskens (2010): "Simultaneous Confidence Bands for Penalized Spline Estimators," *J. of Am. Stat. Assoc.*, 105, 852–863.

Parzen, E. (1979): "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–121.

Pfanzagl, J. (1976): "Investigating the Quantile of an Unknown Distribution," in *Contributions to Applied Statistics: Dedicated to A. Linder*, ed. by W. J. Ziegler. Birkhäuser Verlag.

Reich, B., M. Fuentes, and D. Dunson (2011): "Bayesian Spatial Quantile Regression," *J. of Am Stat. Assoc*, 106, 6–20.

Reiss, R. D. (1980): "Estimation of Quantiles in Certain Nonparametric Models," *Ann. Statist.*, 8, 87–105.

Schabel, S. K., and P. Eilers (2009): "Optimal Expectile Smoothing," *Computational Statistics and Data Analysis*, 53, 4168–4177.

Sheather, S. J., and J. S. Marron (1990): "Kernel Quantile Estimators," *J. of Am. Stat. Assoc.*, 85, 410–416.

Sims, C. (2012): "Robins-Wasserman, Round-N," `http://sims.princeton.edu/yftp/WassermanExmpl/WassermanR4a.pdf`.

Wang, H., D. Li, and X. He (2012): "Estimation of High Conditional Quantiles for Heavy-Tailed Distributions," *J. of Am. Stat. Assoc.*, 107, 1453–1464.

Wei, Y., A. Pere, R. Koenker, and X. He (2005): "Quantile Regression for Reference Growth Charts," *Statistics in Medicine*, 25, 1369–1382.